

CH5350: Applied Time-Series Analysis

Models for Linear Non-Stationary Processes

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

Classical Time-Series Model

Once a time-series is realized as predictable, the search for a suitable mathematical model is carried out. Classical approaches in the early days rested on the philosophy that a series is made up of three components

$$\text{Time-Series} = \text{Trend} + \text{Seasonal Component} + \text{Stationary component}$$

The trend and seasonal components could be combined into a single component under the banner of *deterministic* component.

Classical Approach

... contd.

Several efficient non-parametric and semi-parametric methods were subsequently developed to realize such a decomposition. The trend usually contains a polynomial type of trend while the seasonal component captures the periodic characteristics, if any.

Extracting the deterministic portions of a series is not trivial, but can be effectively carried out with suitable **regression, smoothing and filtering operations**.

Note: The seasonal component is usually a deterministic periodic signal, and assumed to be uncorrelated with the non-seasonal component.

Modern Approach

In 1970s, a new approach to modelling the seasonal (including the non-stationary and trend components) was introduced.

Unlike the models based on additive approach, **multiplicative models** were postulated. These are more generic in the nature because they take into account the correlation between seasonal and non-seasonal (stationary) components, and also model the integrating (random walk) effects.

The resulting models are known as **seasonal ARIMA (SARIMA) models**.

Non-stationarities

ARMA models are capable of representing almost all classes of (linear) stationary processes. On the other hand, there are several processes that exhibit non-stationarities, of which there are numerous variants.

We shall focus on two popular types, **mean** and **variance** non-stationary processes.

In addition, the series may contain **periodicities** and **seasonalities**. We shall study such processes in greater detail later.

Mean non-stationarities

Among the mean non-stationary processes, we have again two types:

- ❶ Deterministic trends (polynomial functions of time)
- ❷ Stochastic trends, esp. random walk type (**integrating** effects)

We shall, be largely concerned with **integrating type non-stationarities**, since these are frequently encountered across different disciplines.

Handling non-stationarities

There are primarily two ways of handling non-stationarities in the mean:

- ① Eliminating trends by first fitting polynomial models to the series and working with the residuals (suited for deterministic changes in mean).
- ② Including integrating effects in the model by suitable differencing of the data (works for both classes of trends, but suitable for stochastic changes in mean).

Exogenous effects

If the non-stationarities are due to external signals (exogenous effects), then those effects have to be removed first by fitting a deterministic model between the exogenous signals and then fitting a time-series model to the residuals. Else the deterministic and time-series models can be jointly estimated as well. **This is the crux of system identification.**

Fitting trends

Trends in series can be eliminated by two different approaches:

- 1 By fitting a trend using polynomial fits
- 2 Estimating the trend by the use of suitable filters

Example

Consider developing a model for the following series

$$v[k] = \underbrace{\alpha_0 + \alpha_1 k}_{m[k]} + w[k] \quad \alpha_0, \alpha_1 \in \mathcal{R}$$

where $v[k]$ is a zero-mean stationary process

Then, the linear trend (in time) can be estimated by **fitting a straight line using a least squares method**.

Example

... contd.

Alternatively,

$$E(v[k]) = E(\alpha_0 + \alpha_1 k + w[k]) + E(w[k]) = \alpha_0 + \alpha_1 k$$

Thus, the deterministic component (trend) is the average of the series $v[k]$. To put this observation into practice, we replace the theoretical average with an estimate.

The filtering approach essentially uses this idea to extract the trend.

Estimating trends using filters

1 Smoothing with a **two-sided finite moving average filter**

$$\hat{m}[k] = \frac{1}{2M+1} \sum_{j=-M}^M v[k-j] \quad (1)$$

- ▶ This moving average filter assumes linear trend over the interval $[k-M, k+M]$ and that the average of the remaining terms is close to zero
- ▶ The filter provides us, therefore, with an estimate of the linear trend

Estimating trends using filters

... contd.

- ② A “clever” choice of filter can be used to eliminate polynomial trends. The Spencer 15-point MA filter can be used to estimate polynomial trends of degree 3.

Spencer’s filter coefficients:

$$a_j = 0, |j| > 7, \quad a_j = a_{-j}, |j| \leq 7$$

a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7
$74/320$	$67/320$	$46/320$	$21/320$	$3/320$	$-5/320$	$-6/320$	$-3/320$

Estimating trends using filters

... contd.

- ③ **Exponential smoothing of data to estimate the trend $m[k]$**

$$\hat{m}[k] = \alpha v[k] + (1 - \alpha)\hat{m}[k - 1], \quad k = 2, \dots, n$$

$$\hat{m}[1] = v[1]$$

The choice of α has to be fine tuned according to the trend.

- ④ **Smoothing by elimination of high-frequency components of the series**

- ▶ A spectral analysis of the data can be carried out to determine the cut-off frequency

Note: When seasonality/periodicity is present, an additional procedure has to be followed that eliminates seasonality prior to trend estimation.

Method of differencing (high-pass filtering)

Consider the same series as earlier

$$v[k] = \underbrace{\alpha_0 + \alpha_1 k}_{m[k]} + w[k] \quad \alpha_0, \alpha_1 \in \mathcal{R}$$

Construct the differenced series

$$v[k] - v[k-1] = (1 - q^{-1})v[k] = \alpha_1 + (1 - q^{-1})w[k]$$

Introducing $\nabla = 1 - q^{-1}$, we can thus observe that the differenced series $\nabla v[k]$ is a non-zero mean stationary process

Method of differencing (high-pass filtering) ... contd.

It is easy to observe that polynomials of d^{th} degree can be eliminated by differencing d times, i.e.,

Differenced series $\nabla^d v[k]$ is free of all polynomial trends up to degree d

However, a single degree of **differencing** introduces a **non-invertible zero** in the differenced series!

Note: The operator ∇^d should not be confused with the other operator $\nabla_d = 1 - q^{-d}$ that is often used to eliminate seasonal effects.

Method of differencing vs. polynomial fits

- ▶ **Filtering approach:** Estimates the trend followed by the development of an ARMA model for the stationary residuals
- ▶ **Differencing approach:** In contrast, eliminates the trend implicitly and fits an ARMA model.
 - ▶ Forces one of the poles to sit on the unit circle.
 - ▶ Introduces a non-invertible zero in the MA term of the stationary component (not desirable).

Fitting polynomials and filtering

- ▶ Removal of linear or polynomial trends by curve fitting seems to be the preferable way of detrending. However, it is not without its shortcomings.
 - ▶ **Major difficulty:** Inability to distinguish between trends and incomplete cycles of very low frequencies (Granger, 1966; Granger and Hatanaka, 1964).
- ▶ **Alternative:** Use MA and exponential filters to extract trends (Brockwell, 2002; Chatfield, 2004)).

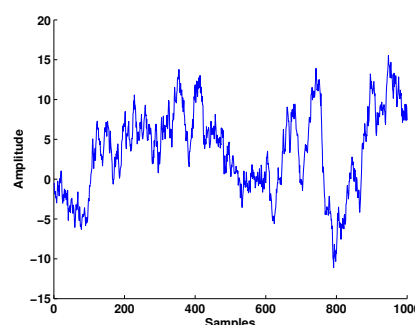
Integrating processes

A pure integrating type stochastic process is described by

$$v[k] = v[k - 1] + e[k] \quad (2)$$

where $e[k]$ is the usual WN.

- For the process (2) above, the series $\nabla v[k] = (1 - q^{-1})v[k] = e[k]$ is indeed stationary.



Sample path of an I(1) process

Integrating type process

... contd.

The integrating process in (2) is also an AR(1) process, but with the pole on the unit circle.

An alternative representation of (2) is

$$v[k] = \sum_{j=0}^{\infty} e[k - j] \quad (3)$$

Thus, this is an MA process of infinite-order. The coefficients of this model, expectantly, do not satisfy the condition of absolute convergence.

Note: The representation in (3) also explains the name for the process.

ARIMA models

In a more general scenario, the differenced series can belong to a correlated stationary process, which can be given an ARMA representation. This is a standard way of handling non-stationary processes of the integrating type.

The resulting representation gives rise to an ARIMA(P, d, M) model:

$$\nabla^d v[k] = \frac{C(q^{-1})}{D(q^{-1})} e[k] \quad (4)$$

where

$$\nabla^d v[k] = (1 - q^{-1})^d v[k] \quad (5)$$

ARIMA models . . . contd.

For the *original* series, the transfer function is therefore

$$H(q^{-1}) = \frac{1 + \sum_{i=1}^M c_i q^{-i}}{(1 - q^{-1})^d (1 + \sum_{j=1}^P d_j q^{-j})} \quad (6)$$

- ▶ The quantities P and M have their usual meaning as in ARMA models
- ▶ The parameter d refers to the order of integrating effect
- ▶ Best remembered as an ARMA model on a series that is differenced d times
- ▶ ARIMA models necessarily have d poles on the unit circle

A more general ARIMA representation is

$$\nabla^d v[k] = \mu_d + \frac{C(q^{-1})}{D(q^{-1})} e[k] \quad (7)$$

- ▶ The constant μ_d can originate from the presence of a *polynomial trend* of the form $\sum_{i=0}^d \mu_i k^i$ in $v[k]$.
- ▶ Thus, when an *ARIMA model with a constant* is fit to a given series, it is understood implicitly that $v[k]$ has a polynomial trend of degree d with or without integrating effects.

We shall mostly work with (4)

Detecting integrating effects via sample ACF / PACF

ACF estimates exhibit a very slow decay for integrating type process.

This characteristic can be understood by examining the theoretical ACF of an AR(1) process,

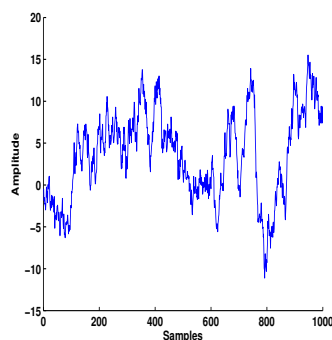
$$\rho[l] = (-d_1)^{|l|} \quad (8)$$

As d_1 approaches unity, the decay is very slow. In the limiting case $d_1 \rightarrow 1$, the ACF approaches a constant.

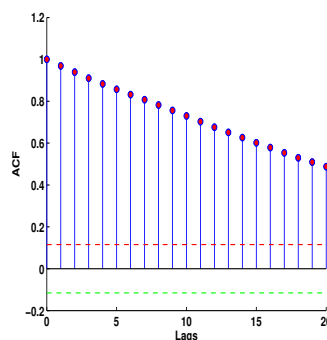
ACF and PACF of an integrating process

- ▶ The exact behavior of the estimates is more complicated, but it tends to have the characteristics of the theoretical ACF for large samples.
- ▶ It follows that the PACF of this process has a single non-zero coefficient at lag $l = 1$ with a value tending to unity.

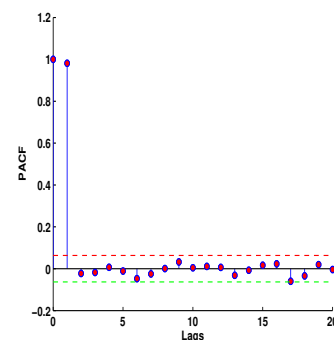
Example



(a) Integrating series



(b) Sample ACF of I(1)



(c) Sample PACf of I(1)

The ACF and PACF estimates agree with the theoretical observations made earlier.

Note: The slow decay in the sample ACF is also the characteristic of a trend-stationary process.

Differencing: Remarks

The differencing operation truly serves its purpose only when the process has a pole location exactly at unity.

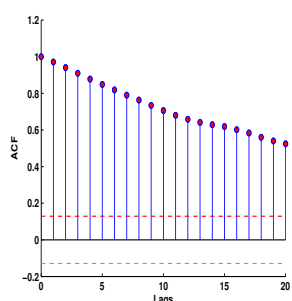
This requirement cannot be expected to hold for a general process.

Is differencing a series worth the risk?

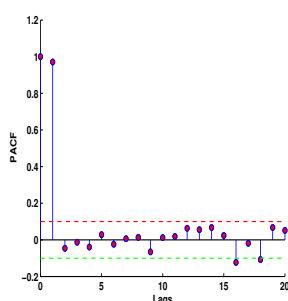
- ▶ Estimation of “stationary” poles, but close to unit circle, can result in models with confidence regions containing non-stationary models. By forcing the pole to the unit circle a priori, this situation is avoided.
- ▶ Side effect: a stationary process with slowly decaying process acquires a non-stationary representation.

Case for differencing: Example

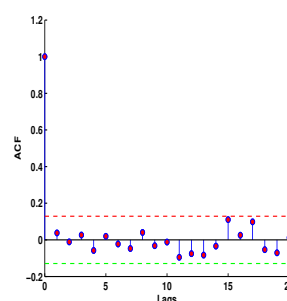
Consider $N = 400$ samples of a stationary process: $H(q^{-1}) = \frac{1}{1 - 0.96q^{-1}}$



(a) ACF of series



(b) PACF of series



(c) PACF of diff. series

The ACF shows a slow decay indicating possible random walk behavior, as confirmed by the PACF.

Example

... contd.

Suppose we fit an AR(1) model to the series,

$$\hat{H}(q^{-1}) = \frac{1}{1 - \underset{0.011}{0.9745}q^{-1}} \quad (9)$$

The pole location of the estimated (nominal) model is at 0.97, which is close to but within the unit circle.

Example

... contd.

- ▶ However, the 99% confidence interval for the parameter d_1 is given by -0.9745 ± 0.03 (we shall learn later how to compute CI).
- ▶ Thus, one of the possible models in the model set has a pole location outside the unit circle. Clearly, this does not agree with the observed series.
- ▶ On the other hand, the I(1) model is better suited for the series since at least it is not explosive, if not stationary.

Overdifferencing

Suppose that the series, or the differenced series $v[k]$, is white. Differencing $v[k]$ yields,

$$\nabla v[k] = v_d[k] = e[k] - e[k-1] = (1 - q^{-1})e[k] \quad (10)$$

A spurious correlation is thus observed in $v_d[k]$ at lag $l = 1$. Once again, we have that an artificial non-invertible zero is introduced in the differenced series.

Use the differencing approach cautiously and perform appropriate tests before developing an ARIMA model.

Tests for overdifferencing

Simple test for overdifferencing

Variance test

When $\text{var}(\nabla^d(v[k])) > \text{var}(v[k])$, overdifferencing has occurred

It is a conservative approach for correlated series.

Unit root in the MA polynomial

The basic idea is that an overdifferenced series contains a zero at unity (or several zeros). See (Brockwell, 2002, Chapter 6) for more details.

Tests for unit roots (in the AR polynomial)

A few different tests for the presence of poles at unity are available:

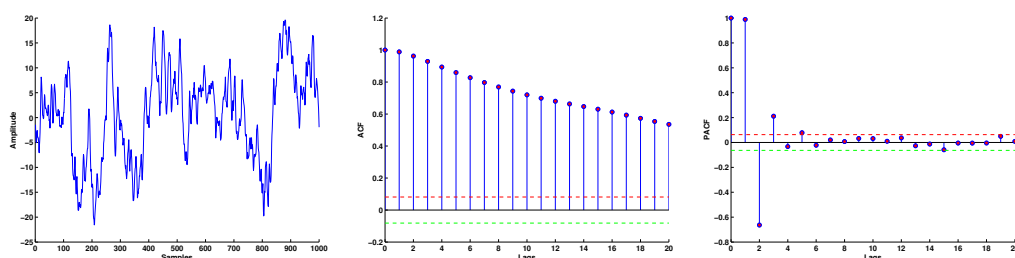
- 1 Augmented Dickey-Fuller (ADF) test (Said and Dickey, 1984)
- 2 Philip-Perron (PP) test (Phillips and Perron, 1988)
- 3 Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test (Kwiatkowski et al., 1992)

Important: Pay attention to the null and alternate hypotheses in each of these tests!

More efficient unit root tests are available. See Elliot, Rothenberg, and Stock, (1996) and Ng and Perron, (2001).

Estimating an ARIMA model

Simulated series, ACF and PACF shown below. Goal is to estimate a suitable ARIMA model.



(a) Simulated series

(b) ACF of the series

(c) PACF of the series

ACF and PACF suggest presence of integrating effect(s). Nevertheless, we explore both routes, i.e., fitting an ARMA and ARIMA model.

ARIMA Modelling: Fitting an ARMA model

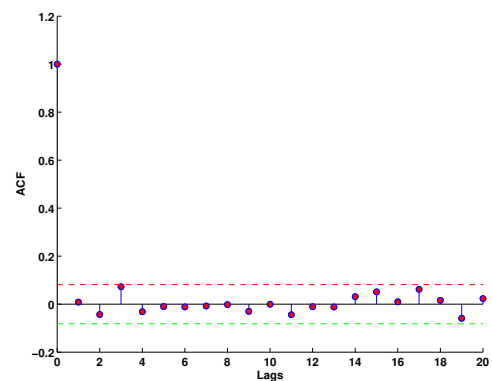
First possibility is to build a model for the series as is, based on the signatures shown by the PACF plot, which suggests an AR(3) model.

The estimated AR(3) model is:

$$\hat{H}_1(q^{-1}) = \frac{1}{1 - \underset{(\pm 0.03)}{1.844}q^{-1} + \underset{(\pm 0.06)}{1.082}q^{-2} - \underset{(\pm 0.03)}{0.2249}q^{-3}}; \quad \hat{\sigma}_e^2 = 0.9617 \quad (11)$$

ARIMA modelling: Analysis of the ARMA model

- ▶ Model passes the residuals test as shown in the adjacent figure.
- ▶ Parameter estimates are reliable (low errors).

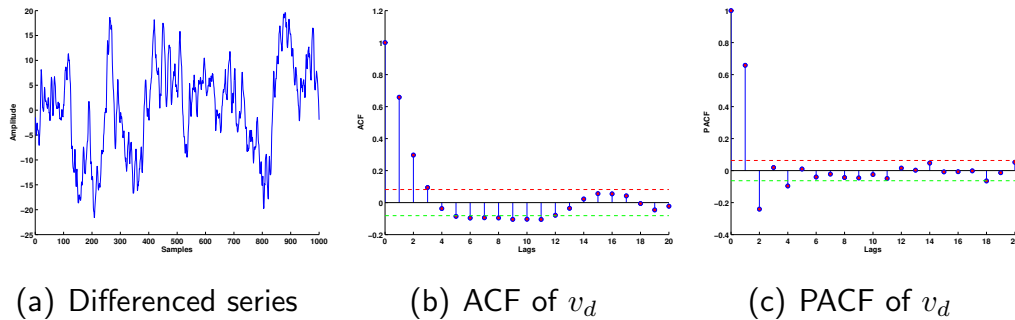


ACF of AR(3) (original series) residuals

Thus, this model is satisfactory in all respects.

ARIMA modelling example: Differencing the series

The differenced series $v_d[k] = \nabla v[k]$, its ACF and PACF are shown below.



The PACF suggests an AR(2) model for $v_d[k]$. For comparison purposes, we estimate an ARMA(1,1) model, which is also characterized by two parameters.

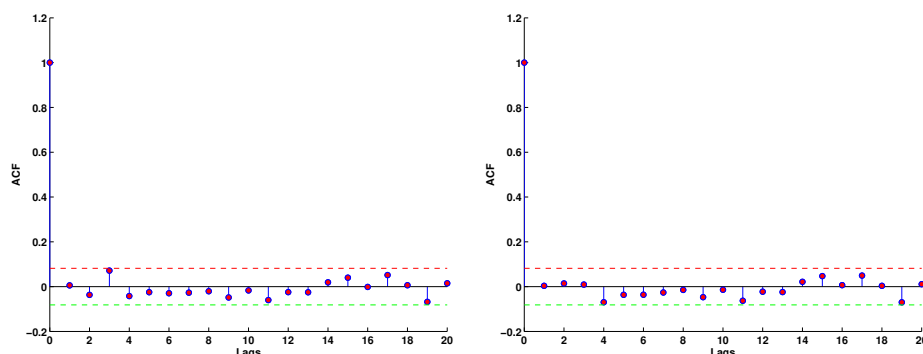
ARIMA Modelling: Differencing approach

The estimated AR(2) and the ARMA(1,1) models are, respectively:

$$\hat{H}(q^{-1}) = \frac{1}{1 - \underset{(\pm 0.03)}{0.821}q^{-1} + \underset{(\pm 0.03)}{0.243}q^{-2}}, \hat{\sigma}_e^2 = 0.9769$$

$$\hat{H}(q^{-1}) = \frac{\underset{(\pm 0.04)}{1 + 0.364}q^{-1}}{1 - \underset{(\pm 0.04)}{0.467}q^{-1}}, \hat{\sigma}_e^2 = 0.9814$$

ARIMA Modelling: Differencing approach ... contd.



(d) ACF of AR(2) (diff. series) residuals
(e) ACF of ARMA(1,1) (diff. series) residuals

ARIMA modelling example ... contd.

- ▶ Both models pass the whiteness test and have reliable parameter estimates.
- ▶ However, we shall accept the AR(2) model because it has unique estimates.

Thus, a suitable ARIMA model for the series is:

$$\hat{H}_2(q^{-1}) = \frac{1}{(1 - q^{-1})(1 - \underset{(\pm 0.03)}{0.821}q^{-1} + \underset{(\pm 0.03)}{0.243}q^{-2})} \quad (12)$$

ARIMA modelling example

... contd.

- ▶ Incidentally the model in (11) for the original (un-differenced) series (11) and the one in (12) for the differenced series are of the same order. The distinguishing feature is the pole locations of these two models, as given below

$$\text{Poles}(\hat{H}_1) : \quad 0.9625, 0.4185 \pm j0.2344$$

$$\text{Poles}(\hat{H}_2) : \quad 1, 0.4104 \pm j0.2732$$

- ▶ The *nominal* poles of \hat{H}_1 are all stable, thus corresponding to a stationary model whereas one of the nominal poles of \hat{H}_2 is on the unit circle. *Which model is preferable?*

ARIMA modelling example

... contd.

The deciding factor is the **confidence region** for each of those models.

- ▶ One of the possible models in the model set of \hat{H}_1 has poles located at $1.0324, 0.4283 \pm j0.3438$ (the model with coefficients $\hat{d}_i + 3\hat{\sigma}_{d_i}$). An explosive model appears in the confidence region, which is not acceptable.
- ▶ The model set associated with \hat{H}_2 does not possess this shortcoming.

In light of these arguments, $\hat{H}_2(q^{-1})$ is selected as the suitable model.

The underlying process

Data generating model:

$$H(q^{-1}) = \frac{1 + 0.3q^{-1}}{(1 - 0.97q^{-1})(1 - 0.54q^{-1})}$$

Variance non-stationarities

Variance non-stationarities (**heteroskedasticity**) are of different kinds.

- ▶ The ARIMA model can handle series with variance that is proportional to time.
- ▶ There are, however, other types of variance non-stationarities:
 - ▶ For example, variance can change as a deterministic function of the mean.
 - ▶ Alternatively, it could be a complicated function of the series.

Approaches to handle heteroskedasticity include building ARMA models on variance stabilizing transformed series or to use the more versatile ARCH and GARCH models.

Variance stabilizing transformation

Several processes exhibit a property whose variance changes with the level.

$$\sigma_k^2 = Ch(\mu_k) \quad (13)$$

Objective: Find a transformation $g(y[k])$ s.t. the transformed series has a constant variance.

Use a first-order approximation for $g(\cdot)$ using Taylor's-series expansion

$$g(y[k]) \approx g(\mu_k) + (y[k] - \mu_k)g'(\mu_k) \quad (14)$$

and demand that σ_k^2 be a constant.

Solution

The solution is given by

$$g'(\mu_k) = \frac{1}{\sqrt{h(\mu_k)}} \quad (15)$$

A more general transformation was suggested by Box and Cox (1964):

$$y_\lambda[k] = \begin{cases} \frac{(y[k])^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(y[k]), & \lambda = 0 \end{cases} \quad (16)$$

where λ is the **transformation parameter** - user-specified or optimized by an algorithm.

Box-Cox transformation

λ	Transformation
-1.0	$1/v[k]$
-0.5	$1/\sqrt{v[k]}$
0.0	$\ln(v[k])$
0.5	$\sqrt{v[k]}$
1.0	$v[k]$ (no transformation)

- ▶ Valid only for **positive-valued series**.
- ▶ Transformations can aid in even improving the approximation of non-Gaussian distribution with a Gaussian one.
- ▶ However, they can also result in violations of Gaussian distribution and other assumptions!

Summary

- ▶ Trend-type (with seasonal) non-stationarities are handled by explicitly fitting a deterministic component (either through polynomial fit or by applying a suitable filter) followed by an ARMA modelling of the residual.
- ▶ Integrating effects are modelled by fitting ARIMA models, which is equivalent to fitting an ARMA model to the differenced series.
- ▶ Overdifferencing can introduce non-invertible zeros in the model.
- ▶ Integrating effects (or slowly decaying nature of stationary processes) are detected by slowly decaying ACFs, or more rigorously by unit root tests.
- ▶ Heteroskedastic series are handled by either variance stabilizing transformations or through ARCH / GARCH models.